REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01/22/2015	Final	04/15/2011 - 10/14/14

4. TITLE AND SUBTITLE

Photon-limited sensing and surveillance

5a. CONTRACT NUMBER
n/a

5b. GRANT NUMBER
FA9550-11-1-0028

5c. PROGRAM ELEMENT NUMBER
n/a

6. AUTHOR(S)
Robert Calderbank, Ph.D.
Rebecca Willet, Ph.D.

5d. PROJECT NUMBER
n/a

5e. TASK NUMBER
n/a

5f. WORK UNIT NUMBER
n/a

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Duke University
Campus Box 90984
315 Gross Hall
140 Science Drive

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
USAF,AFRL, DUNS 143574726
875 North Randolph Street, Room 3112
Arlington VA 22204
Yvonne M. Mason 703-696-9518

10. SPONSOR/MONITOR'S ACRONYM(S)
USAF, AFRL

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
n/a

12. DISTRIBUTION / AVAILABILITY STATEMENT

A- Approved for Public Release

13. SUPPLEMENTARY NOTES

n/a

14. ABSTRACT

The goal of this work was to learn and exploit unknown spatio-temporal structure in online photon-limited sensing and surveillance data. Photon-limited imaging arises in a wide variety of applications of interest to the Air Force, including night vision, space weather, imaging through fog, and spectral imaging. The photon-limited video reconstruction problem is particularly challenging because (a) the limited number of available photons introduces intensity-dependent Poisson statistics which require specialized algorithms and analysis for optimal performance, (b) vast quantities of video data will be collected sequentially, necessitating fast online algorithms, and (c) unknown and changing environmental dynamics preclude classical methods based on known dynamical models. Many current systems sidestep photon limitations by artificially restricting the frame rate and resolution of the video, but sophisticated statistical methods allow dramatic increases in resolution and improved object identification and detection capabilities.

15. SUBJECT TERMS

Night vision, photon-limited imaging, sparse coding, collaborative filtering, online learning, compressed sensing, Poisson noise

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	SAR		Robert Calderbank, Rebecca Willet
U	U	U			**19b. TELEPHONE NUMBER** *(include area code)* 919-660-5286

AFRL-OSR-VA-TR-2015-0066

Photon-limited Sensing and Surveillance

Robert Calderbank
DUKE UNIVERSITY

01/29/2015
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTC
Arlington, Virginia 22203
Air Force Materiel Command

Photon-limited sensing and surveillance: Final Report for AFOSR grant FA9550-11-1-0028

Duke University PI Robert Calderbank
University of Wisconsin-Madison PI Rebecca Willett

January 2, 2015

Contents

1 Summary of program objectives and outcomes

The goal of this work was to learn and exploit unknown spatio-temporal structure in online photon-limited sensing and surveillance data. Photon-limited imaging arises in a wide variety of applications of interest to the Air Force, including night vision, space weather, imaging through fog, and spectral imaging. The photon-limited video reconstruction problem is particularly challenging because (a) the limited number of available photons introduces intensity-dependent Poisson statistics which require specialized algorithms and analysis for optimal performance, (b) vast quantities of video data will be collected sequentially, necessitating fast online algorithms, and (c) unknown and changing environmental dynamics preclude classical methods based on known dynamical models. Many current systems sidestep photon limitations by artificially restricting the frame rate and resolution of the video, but sophisticated statistical methods allow dramatic increases in resolution and improved object identification and detection capabilities.

We addressed these challenges by developing new tools for learning and exploiting low-dimensional signal structured, including sparsity and low-rank structure, from photon-limited data. In addition, we developed novel online learning methods that would allow large-scale photon-limited video data to be analyzed as it was collected (as opposed to forensic analysis with a considerable time delay). More specifically, there were four main outcomes from this work:

- **Improved understanding of the fundmental limitations of compressed sensing (CS) for photon-limited imaging.** Several engineers and scientists from the optics and signal processing communities have suggested that we design novel cameras for photon-limited settings based on the principles of CS. Most prior theoretical results in compressed sensing and related inverse problems apply to idealized settings where the noise is i.i.d., and do not account for signal-dependent noise and physical sensing constraints. Prior results on Poisson compressed sensing with signal-dependent noise and physical constraints in [16] provided upper bounds on mean squared error performance for a specific class of estimators. However, it was unknown whether those bounds were tight or if other estimators could achieve significantly better performance. Our work provided minimax lower bounds on mean-squared error for sparse Poisson inverse problems under physical constraints, and *demonstrate the CS is not a viable paradigm for photon-limited sensing and surveillance*. For additional details, see

  > X. Jiang, G. Raskutti, and R. Willett, "Minimax optimal rates for Poisson inverse problems with physical constraints", accepted to *IEEE Transactions on Information Theory*, arXiv:1403:6532, 2014.

- **Novel method for photon-limited signal denoising which represents the current state-of-the-art.** We have developed denoising algorithms for photon-limited images which combine elements of dictionary learning and sparse representations for image patches [17]. Our preliminary method employs both an adaptation of Principal Component Analysis (PCA) for Poisson noise and our sparsity regularized convex optimization algorithms for photon-limited images. A comprehensive empirical evaluation of the proposed method reveals that, despite its simplicity, PCA-flavored denoising appears to be highly competitive in very low light regimes, as depicted in Figure 2. In this figure, we compare with BM3D, widely considered to be the current state-of-the-art for image denoising, and other widely-used Poisson image denoising methods. For more details, see

  > J. Salmon, Z. Harmany, C. Deledalle, and R. Willett, "Poisson noise reduction with non-local PCA," *Journal of Mathematical Imaging and Vision*, vol. 48, no. 2, pp. 279-294, arXiv:1206:0338, 2014.

- **Reparameterizations of photon-limited images.** Most photon-limited image reconstruction methods optimize a regularized negative Poisson log likelihood to estimate the underlying image intensity.

However, at very low intensities, this focus on the image intensity leads to several technical challenges related to efficient optimization tools, cross-validation techniques, and empirical performance. To combat this, we introduced a novel approach for Poisson image reconstruction that adapts to the signal intensity level through a hybrid objective function with useful properties. This method performs well visually and empirically, and outperforms prior models in terms of RMSE. Our method is also amenable to cross-validation for selecting model parameters accurately.

> A. K. Oh, Z. T. Harmany, and R. M. Willett, "To *e* or not to *e* in Poisson image reconstruction", *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2014. Received award as "Top 10% Paper".

A journal version of this work is in progress.

- **New approaches for online photon-limited video reconstruction and analysis.** With streaming photon-limited video, it is possible to compute more accurate reconstructions by exploiting not only spatial structured, as describe above, but also temporal structure. Such reconstructions are essential for subsequent analysis, such as foreground and background separation. However, standard stochastic filtering methods (like Kalman or particle filters) are ill-suited for this regime. We developed novel online learning methods which are capable not only of accounting for photon limitations, but also learn and exploit the underlying scene dynamics. The supporting theory is a unique contribution to the online learning literature, while the algorithms are fast and produce state-of-the-art reconstructions.

  > E. Hall and R. Willett, "Foreground and background reconstruction in Poisson video", *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2013.

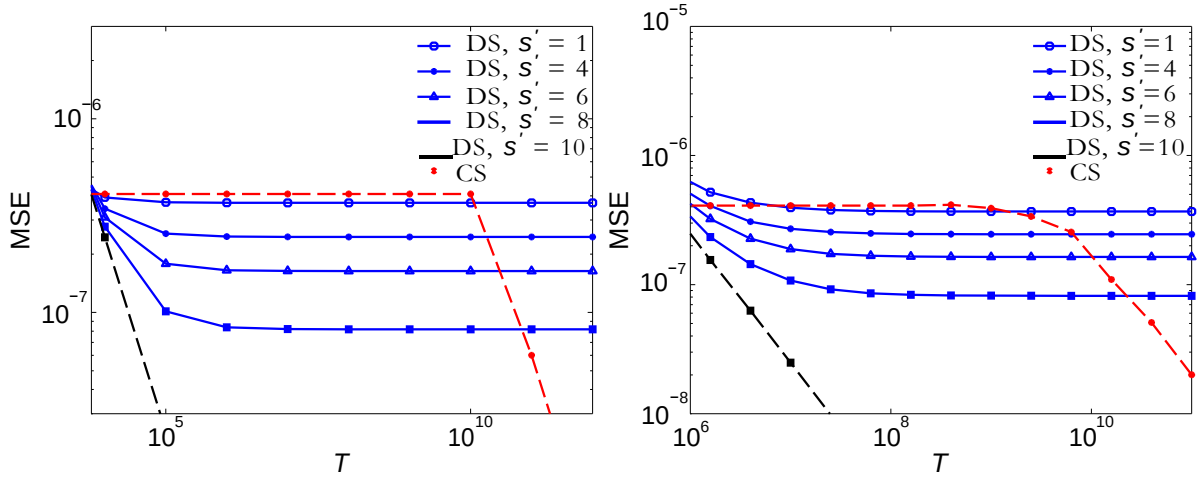
  A journal version of this work is in progress and is detailed below.

2 Relationship between program outcomes and previous state-of-the-art

The four major outcomes all represent advances on the previous state-of-the-art.

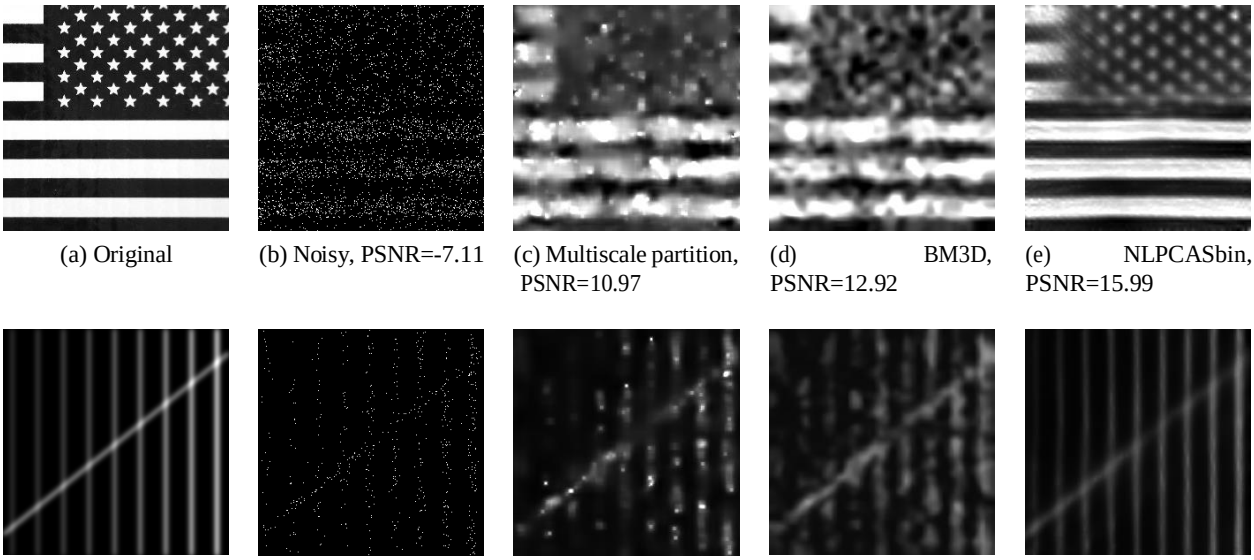
Photon-limited CS. The bounds on compressed sensing for photon-limited imaging are the first known lower bounds for this problem and highlight the practical challenges of using CS for night vision in USAF equipment. An example of this is depicted in Figure 1. Specifically, we consider images which are s-sparse in a wavelet basis, and s^t of the s non-zero wavelet coefficients are at a coarse scale. (In many natural images, s^t is large relative to s.) The conventional wisdom of CS tells us that CS is preferable to simply using a low-resolution imager – *i.e.,* downsampling the scene (DS) – because CS will recover *both* the s^t coarse-scale coefficients and the $s - s^t$ fine-scale coefficients, while a low-resolution imager only allow recovery of the s^t coarse-scale coefficients. Our theory and supporting simulations demonstrate that this **conventional wisdom is incorrect in high-noise, photon-limited regimes**, and DS can significantly outperform CS.

Non-local PCA for photon-limited image estimation. As described in the previous section, our method for photon-limited image estimation, which leverages ideas from dictionary learning and sparse recovery methods, represents the current state-of-the-art in photon-limited imaging. Empirical results shown in Figure 2 highlight the improvements possible via the work supported by this grant.



(a) Theoretical rates (b) Empirical rates

Figure 1: Theoretical and empirical rates of downsampling and compressed sensing methods. Plots correspond to imaging a scene with 2048 pixels using 512 measurements when the scene as $s = 10$ non-zero wavelet coefficients. s^t is the number of coarse-scale nonzero coefficients which are directly measured by the proposed downsampling scheme. We see that at low-intensities, downsampling can yield much lower MSEs, but after the intensity exceeds a critical threshold, compressed sensing methods are able to estimate all nonzero coefficients accurately and the MSE is better than for downsampling schemes. This effect is predicted by our theory.

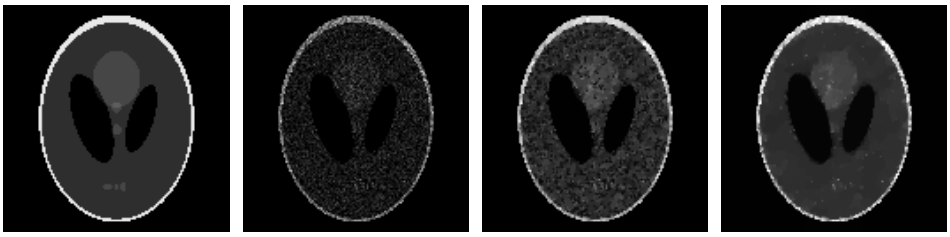


(a) Original (b) Noisy, PSNR=-7.11 (c) Multiscale partition, PSNR=10.97 (d) BM3D, PSNR=12.92 (e) NLPCASbin, PSNR=15.99

(f) Original (g) Noisy, PSNR=0.31 (h) haarTIApprox, PSNR=18.69 (i) BM3D, PSNR=19.30 (j) NLPCASbin, PSNR=23.27

Figure 2: Simulated images (flag in top row, ridges in bottom row) corrupted with Poisson noise with peak intensity of 0.1. Our method is NLPCASbin, and its result is a notable improvement over the previous state of the art.

Reparameterizations of photon-limited images. The reparameterization of photon-limited images allows fast and accurate recovery with empirical advantages over classical reconstruction approaches. This



(a) Truth scene intensity

(b) Photon-limited observations

(c) Classical method recovery using total-variation regularized Poisson log-likelihood

(d) Recovery using total-variation regularization with our reparameterization

Figure 3: Denoising results using our reparameterization method.

is illustrated in Figure 3, where we recover an image from photon-limited observations using (c) classical total-variation regularization of a negative Poisson log-likelihood and (d) total-variation regularization applied to our our reparameterization of the Poisson log-likelihood. The primary advance over previous work is a new perspective on how regularizers can and should be selected in photon-limited imaging. The classical approach is to determine from the outset that the scene will be parameterized by a linear function of its pixel intensity values, and then choose a convex regularization function that will facilitate using off-the-shelf convex optimization tools to compute an image estimate. In contrast, our approach is far more flexible in that regularizers can be applied to non-linear transformation of the image (*e.g.*, its logarithm), giving us access to a much wider array of potential regularizers that can be used with convex optimization tools. These regularizers can then be used to improve photon-limited image reconstruction.

Online reconstruction of streaming photon-limited video. Our methods for reconstructing photon-limited video, particularly separating moving foreground and background from low-SNR data, is novel and an advance over the previous state-of-the-art in several respect. Classical stochastic filtering methods such as Kalman or particle filters or Bayesian updates [2] readily exploit dynamical models for effective prediction and tracking performance. However, classical methods are also limited in their applicability because (a) they typically assume an accurate, fully known dynamical model and (b) they rely on strong assumptions regarding a generative model of the observations. Some techniques have been proposed to learn the dynamics [20, 21], but the underlying model still places heavy restrictions on the nature of the data. The Kalman filter relies on Gaussian noise models, and particle filters exhibit particle degeneracy, making them difficult to use in practical settings.

A contrasting class of prediction methods is based on an "individual sequence" or "universal prediction" [13] perspective; these strive to perform provably well on any individual observation sequence. In particular, online convex programming methods [4, 7, 14, 23] rely on the gradient of the instantaneous loss of a predictor to update the prediction for the next data point. The aim of these methods is to ensure that the per-round performance approaches that of the best *offline* method with access to the entire data sequence. This approach allows one to sidestep challenging issues associated with statistically dependent or non-stochastic observations, misspecified generative models, and corrupted observations. This framework is limited as well, however, because performance bounds are typically relative to either static or piecewise constant comparators and do not adequately reflect adaptivity to a dynamic environment.

Our approach is novel framework for prediction in the individual sequence setting which incorporates dynamical models effectively a novel combination of state updating from stochastic filter theory and online convex optimization from universal prediction. This framework, and its application to photon-limited video reconstruction, is detailed in the next section.

3 Online Foreground and Background Reconstruction in Poisson Video

Many imaging applications such as night vision, infrared imaging, and certain astronomical imaging systems are characterized by limited amounts of available light. In these and other settings, the goal is to reconstruct spatially distributed and dynamic phenomena from data collected by counting discrete independent events, such as photons hitting a detector. More specifically, we can model our observations at time t as

$$y_t \sim \text{Poisson}(\lambda_t), \tag{1}$$

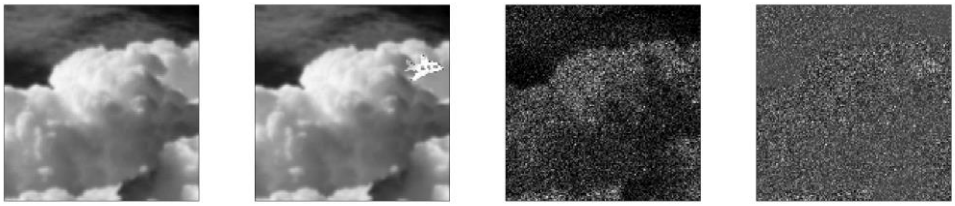
where $y_t \in \mathbf{Z}_+^n$ is the vector of photon counts across n detectors and $\lambda_t \in \mathbf{R}_+^n$ is the intensity of interest (i.e., the n-pixel scene) [18].

We are interested in the case where λ_t has two components: a dynamic foreground φ_t which occupies a relatively small portion of the scene, and a static or slowly-varying background β_t, so that

$$\lambda_t = \varphi_t + \beta_t.$$

The goal here is to recover an accurate estimate of φ_t and β_t from y_t, especially when the photon counts are very low and when the underlying dynamics are unknown.

There exists a rich literature on image estimation and background subtraction methods, and a wide variety of effective tools in high SNR regimes. For instance, a common method for object tracking is to form an estimate of the background scene, and subtract this from the observation to get an estimate for the foreground [15]. Many of these methods make the assumption that the observed pixel values are the true scene corrupted with white Gaussian distributed around the true, slowly varying background mean [19], which is untrue both by the Poisson observation model and settings with dynamic backgrounds. Alternatively, another technique is to learn and track a low-dimensional subspace representation of the background [3]. While such a method can be modified for the Poisson setting, simply subtracting this background estimate from the observations will still not yield an accurate foreground estimate in the low-light setting. In fact, even if the background were known exactly, subtraction will not give a very accurate estimate of the foreground, as shown in Figure 4.



(a) Known Background (b) True Scene (c) Poisson Observation (d) Observation minus *true* background

Figure 4: Challenges of background subtractions for photon-limited video. The background (a) and a foreground object in the top right corner form the true scene (b). Poisson observations are then collected from the true scene (c). Even if the background was known exactly and subtracted from the observations, the resulting image (d) is still very noisy, making accurate inference about foreground objects challenging.

The photon-limited image estimation problem is particularly challenging because it introduces intensity-dependent Poisson statistics which require specialized algorithms and analysis for optimal performance. Simply transforming Poisson data to produce data with approximately Gaussian noise (via, for instance, the

variance stabilizing Anscombe transform [1, 12] or Fisz transform [9, 10]) can be effective when the number of counts is sufficiently high [6, 22]. However, applying these methods to foreground estimation is a difficult problem due to the non-linearities induced by the transforms. Specifically, these tools may make it possible to estimate $\lambda_{t,i}$ effectively, but the inverse problem of estimating φ_t and β_t is significantly more challenging because of the nonlinear relationship between the unknowns and variance stabilized observations.

In addition, the dynamic setting presents significant opportunity for improved photon-limited surveillance. Consider the case in which the temporal dynamics are known exactly. For the Gaussian noise setting, the Kalman filter has proved enormously effective. The known dynamics can effectively act as a prior probability model for the scene at time t, and once y_t has been observed, this prior knowledge can dramatically improve reconstruction accuracy even when the number of available photons is small.

Generalizations of this approach to Poisson noise are possible with particle filters [2], but particle degeneracy is a major practical challenge. Furthermore, classical stochastic filtering methods typically assume an accurate, fully known dynamical model; if a dynamical model is learned from data, it is typically assumed not to change over time.

We present an online method which estimates the underlying, time-varying dynamical model, and uses this estimate to generate online estimates of the foreground and background video sequences. Our approach is based on recent advances in online convex programming and online learning [4, 7, 14, 23]. In particular, we use a variant of Mirror Descent [4, 14] which incorporates dynamical model estimates [11].

3.1 Problem Formulation

We model the data as Poisson observations of a scene which is mostly background with some dynamic foreground. In order to distinguish foreground from background, we assume that the two have discernibly different underlying dynamics, and that the foreground obscures only a small part of the background. We denote the observation at time t as y_t, the background as β_t and the foreground as φ_t. Therefore the probability density function of the observation is given as:

$$p(y_t/\varphi_t,\beta_t) = \prod_{i=1}^{m} \frac{(\varphi_{t,i}+\beta_{t,i})^{y_{t,i}}}{y_{t,i}!} \exp\left(-(\varphi_{t,i}+\beta_{t,i})\right). \tag{2}$$

Here, t indicates time index, and i indicates pixel location. Notice that this model assumes that the observed scene is the superposition of background and foreground at every pixel. In actuality every pixel would either be completely foreground or completely background, but it is difficult to model this explicitly because the locations of the foreground pixels would need to be known exactly *a priori*. Using this model we wish to reconstruct β_t and φ_t as accurately as possible in a time-efficient manner.

3.2 Dynamic Fixed Share algorithm

In order to solve the problems of background and foreground estimation, we will use an algorithm called Dynamic Fixed Share (DFS) [11]. In this section, we describe the DFS method in a general setting, and it's application to background subtraction problems will be described in the next section.

DFS takes in streaming observations and a family of candidate dynamic models $\{\Phi^{(1)},\Phi^{(2)},\ldots\Phi^{(N)}\}$ to produce a sequence of estimates (denoted $\hat{\theta}_t$) with provably low loss. Specifically, at time t we make an observation y_t, and it induces a convex loss function

$$f_t(\theta) = f_t(\theta) + r(\theta),$$

where $f_t(\theta)$ describes how well a candidate estimate θ fits the observation y_t and $r(\theta)$ is a regularization function.

DFS works in two steps, the first being to produce an estimate for each candidate dynamical model $\Phi^{(i)}$ at each time step in the following way:

$$\tilde{\theta}_{t+1} = \arg\min_{\theta \in \Theta} \eta_t \langle \nabla f_t(\hat{\theta}_t), \theta \rangle + \eta_t r(\theta) + D(\theta \| \hat{\theta}_t) \tag{3}$$

$$\hat{\theta}_{t+1} = \Phi^{(i)}(\tilde{\theta}_{t+1}). \tag{4}$$

Here, η_t is a step size parameter and $D(\cdot\|\cdot)$ is a Bregman Divergence. These equations effectively update the previous estimate by taking a step in the direction of the negative gradient of f_t, while also ensuring that the new estimate is well regularized and close to the previous estimate. Once this intermediate estimate is found (3), the dynamical model is applied to get the next estimate (4). The second part of DFS is to produce a single estimate from all of the sub-estimates produced by individual dynamic models. It does this by taking a weighted average of the sub-estimates, with weights based on the accumulated loss of each candidate model.

We characterize the performance of this approach via a *regret bound*, which quantifies the difference between the accumulated loss of our method and the accumulated loss of any comparator sequence θ_t which might be output by a competing, potentially batch, method. It is shown that the estimate produced by the DFS method satisfies the following regret bound:

$$\sum_{t=1}^{T} f_t(\hat{\theta}_t) - \min_{\theta_1,\theta_2,\ldots,\theta_T} \sum_{t=1}^{T} f_t(\theta_t) \leq O\sqrt{T}\,(m+1)(\log(N)+1) + \log \frac{1}{\alpha^m (1-\alpha)^{(T-m-1)}} + \min_{t_2,\ldots,t_{m+1}} \sum_{k=1}^{m+1} \min_{i_k \in \{1,\ldots,N\}} \sum_{t=t_k}^{t_{k+1}-1} \|\theta_{t+1} - \Phi^{(i_k)}(\theta_t)\|_1 ,$$

where N is the number of dynamic models considered, m is the maximum amount of times the optimal dynamic models used to describe the comparator sequence can switch, and α is a parameter used in the algorithm between 0 and 1, which is an estimate on the fraction of times the underlying dynamic model should switch (approximately m/T). The final line of the bound measures how well the comparator sequence, $\theta_1, \theta_2, \ldots, \theta_T$ follows the dynamics on $m+1$ optimally chosen time segments. This variation term finds the best dynamical model in our family and the optimal time points such that the variation term is minimized. This means that if the comparator sequence can be appropriately described as a series of a few subsequences which each closely follow one of the dynamical models , then the regret bound will be low. For more details see [11].

It is important to note that we use the DFS algorithm for the background instead of a moving average:

$$\hat{\beta}_t = \frac{\sum_{s=1}^{t} \alpha^{t-s} y_s}{\sum_{s=1}^{t} \alpha^{t-s}}$$

for some $\alpha \in [0,1]$. This is important because if the background has some dynamic motion, the moving average would perform poorly. If α were set too low, then the background estimate would be heavily corrupted by Poisson noise artifacts. On the other hand, if α were very close to 1, the motion of the background would cause blur in the estimate. Even if α is chosen in between these two extremes, the estimate would not reflect the true background very well as shown in figure 5.

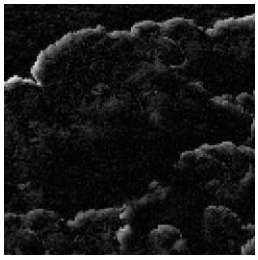


Figure 5: Absolute difference between moving average and true background with $\alpha = .99$. The true background has max value of 5, meaning the errors are relatively large. Notice that this image contains both errors at the leading edge due to motion, and noise errors from the observation model. Both of these errors would adversely affect the foreground estimation performance

3.3 Method

Our first step is to find an estimate the background, so we must first find a loss function for estimating β_t. We use the negative Poisson log-likelihood function of the observation omitting the $y_i!$ term since it is an offset not dependent on β:

$$f_{\beta,t}(\beta) = \sum_{i=1}^{d} \left(\beta_i - y_i \log(\beta_i + \gamma) \right). \tag{5}$$

A small constant, γ is added to ensure numerical stability. Notice that this is the same loss function that would be used if the video sequence was assumed to only have background content.

We then wish to estimate φ_t. We again start by using the negative Poisson log-likelihood as a basis for the loss function for φ_t, but now assume access to an estimate of the background, $\hat{\beta}_t$.

$$-\log(p(y_t/\varphi,\hat{\beta}_t)) = \sum_{i=1}^{d} \left\{ \varphi_i + \hat{\beta}_{t,i} - \log \left\{ \frac{(\varphi_i + \hat{\beta}_{t,i})^{y_{t,i}}}{y_{t,i}!} \right\} \right\} \tag{6}$$

Assuming that the background estimate has already been found, this leads to the following data fit function for the foreground:

$$f_{\varphi,t}(\varphi;\hat{\beta}_t) = \sum_{i=1}^{d} \left\{ \varphi_i - y_{t,i} \log \left\{ \frac{\varphi_i}{\hat{\beta}_{t,i} + \gamma} + 1 \right\} \right\}. \tag{7}$$

This loss function comes from the negative log-likelihood function by subtracting $\sum_{i=1}^{d} \hat{\beta}_{t,i} + \log(y_{t,i}!) - y_{t,i} \log(\hat{\beta}_{t,i})$ which is independent of φ_t. Again, a small positive constant γ is used to ensure numerical stability.

Finally, we include regularization penalties, $r_\beta$and r_φ. For this application, we use a total variation penalty [5, 8], which insures that the estimates are somewhat smooth, as would be expected in natural images. This makes the overall loss functions the following:

$$f_{\beta,t}(\beta) = f_{\beta,t}(\beta) + r_\beta \|\beta\|_{\mathrm{TV}} \tag{8}$$

$$f_{\varphi,t}(\varphi;\beta) = f_{\varphi,t}(\varphi;\beta) + r_\varphi \|\varphi\|_{\mathrm{TV}}. \tag{9}$$

where r_β and r_φ are tradeoff parameters between data fidelity and regularization for the background and foreground respectively. Notice how this process essentially tries to find a coarse estimate for the underlying scene in the background, and then tries to find a foreground which fine tunes this estimate. These loss

Algorithm 1 Background and Foreground Estimation

for $t = 1, \ldots, T$ **do**

Observe y_t

for $i = 1, \ldots N_1$ **do**

$\tilde{w}^{\beta}_{i,t+1} = w^{\beta}_{i,t} \exp(-\eta f_{\beta,t}(\hat{\beta}_{i,t}))$

$w^{\beta}_{i,t+1} = \frac{\lambda}{N_1} \sum_{j=1}^{N_1} \tilde{w}^{\beta}_{j,t+1} + (1-\lambda)\tilde{w}^{\beta}_{i,t+1}$

$\tilde{\beta}_{i,t+1} = \arg\min_{\beta \in B} \eta_t \langle \nabla f_{\beta,t}(\hat{\beta}_{i,t}), \beta \rangle + \tau_\beta \|\beta\|_{\mathrm{TV}} + \ldots$

$\ldots \|\beta - \hat{\beta}_{i,t}\|^2$

$\hat{\beta}_{i,t+1} = \Phi^{(\beta)}_i(\tilde{\beta}_{i,t+1})$

end for

$\beta_{t+1} = \sum_{i=1}^{N_1} \tilde{\beta}_{i,t+1} / \sum_{i=1}^{N_1} w^{\beta}_{i,t+1}$

$\hat{\beta}_{t+1} = \sum_{i=1}^{N_1} w^{\beta}_{i,t+1} \hat{\beta}_{i,t+1} / \sum_{i=1}^{N_1} w^{\beta}_{i,t+1}$

for $k = 1, \ldots, N_2$ **do**

$\tilde{w}^{\varphi}_{k,t+1} = w^{\varphi}_{k,t} \exp(-\eta f_{\varphi,t}(\hat{\varphi}_{k,t}; \tilde{\beta}_{t+1}))$

$w^{\varphi}_{k,t+1} = \frac{\lambda}{N_1} \sum_{j=1}^{N_2} \tilde{w}^{\varphi}_{j,t+1} + (1-\lambda)\tilde{w}^{\varphi}_{k,t+1}$

$\tilde{\varphi}_{k,t+1} = \arg\min_{\varphi \in F} \eta_t \langle \nabla f_{\varphi,t}(\hat{\varphi}_{k,t}; \tilde{\beta}_{t+1}), \varphi \rangle + \ldots$

$\tau_\varphi \|\varphi\|_{\mathrm{TV}} + \|\varphi - \hat{\varphi}_{k,t}\|^2$

$\hat{\varphi}_{k,t+1} = \mathrm{SoftThresh}(\Phi^{(\varphi)}_k(\tilde{\varphi}_{k,t+1}))$

end for

$\varphi_{t+1} = \sum_{k=1}^{N_2} \tilde{\varphi}_{k,t+1} / \sum_{k=1}^{N_2} w^{\varphi}_{k,t+1}$

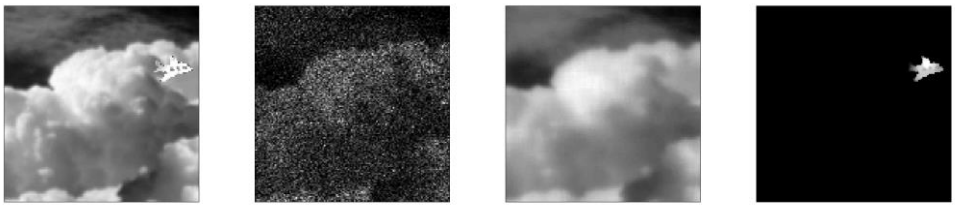
$\hat{\varphi}_{t+1} = \sum_{k=1}^{N_2} w^{\varphi}_{k,t+1} \hat{\varphi}_{k,t+1} / \sum_{k=1}^{N_2} w^{\varphi}_{k,t+1}$

end for

functions, as constructed, are convex which means we can use online convex optimization techniques. We will also use the fact that the background and foreground should have different dynamics to help with separation and reconstruction.

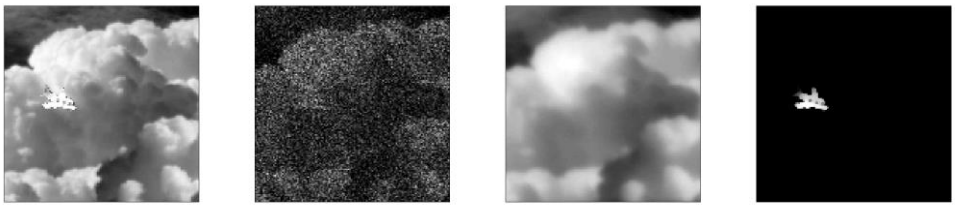
The overall procedure is described in algorithm 1. For both of the inner loops, the minimization was found by using the FISTA algorithm of Beck and Teboulle [5]. Additionally, a small amount of soft thresholding is applied to the foreground estimate at each time step, to ensure that ambiguous areas that could be considered either background or foreground are removed from the foreground estimate. Without this thresholding, these ambiguous areas would appear in the foreground estimate as an underlying haze. It is important to notice for the background and foreground we have two slightly different estimates. The values denoted $\tilde{\beta}_t$ and $\tilde{\varphi}_t$ are the filtering estimates, meaning that they are reconstructions for time t using all the observations up to time t. The values $\hat{\beta}_{t+1}$ and $\hat{\varphi}_{t+1}$ are the prediction values, meaning they use all the data up to time t to predict the observation at time $t+1$.

3.4 Experimental Results



(a) True Scene (b) Poisson Observation (c) Background Estimate (d) Foreground Estimate

Figure 6: Foreground and background reconstruction at $t = 250$. The true image (a) has foreground and background content, and the observations (b) are extremely noisy. We form a background estimate (c) and use it to obtain a foreground estimate (d). Notice the details visible in the foreground estimate such as the windows and tail structure of the plane.



(a) True Scene (b) Poisson Observation (c) Background Estimate (d) Foreground Estimate

Figure 7: Foreground and background reconstruction at $t = 925$. Again notice how the foreground object in (a) is basically imperceptible in the observations (b). By estimating the background (c) an accurate foreground estimate can be constructed (d)

To test this method, we created a data set that featured an object moving across a slowly varying back-

ground in the following way:

$$\varphi_t^* = \Phi_{t-1}^{(\varphi)}(\varphi_{t-1}^*)$$
$$\beta_t^* = \Phi_{t-1}^{(\beta)}(\beta_{t-1}^*)$$
$$x_t = \sum_{i=1}^{d} e_i^T \left[*_{\varphi,t}(i)\varphi_t^* + (1 - *_{\varphi,t}(i))\beta_t^* \right] e_i$$
$$y_t \sim \text{Poisson}(x_t),$$

where $*_{\varphi(t),t}(i)$ is the indicator of pixel i being foreground or not and e_i is the i^{th} standard basis vector . This process shows a foreground object being translated through the function $\Phi_t^{(\varphi)}$ on top of a background image moving with dynamics $\Phi_t^{(\beta)}$. The images were compiled by letting certain pixels be designated as foreground object, and everything else being background. Notice, that the algorithm assumes every pixel is the addition of foreground and background, but the data used is more realistic in that each pixel is either one or the other.

Each image is 150×150, and no pixel has mean value greater than 5, so the video is extremely photon limited. For the background, the true underlying dynamics was a subpixel shift of 1/50th of a pixel to the top left at every time step. For the foreground the true dynamics is a full pixel shift to the top right for the first 500 frames and bottom right for the second 500 frames. The candidate dynamic models used for the background were subpixel shifts of 1/50th of a pixel shift in directions of $k\pi/4$ for $k = 1, 2, \ldots 8$ and stationary ($\Phi^{\beta} = I$). The foreground candidate dynamics were full pixel shifts in the same directions as well as a stationary dynamic.

Figures 6 and 7 show examples of the DFS algorithm taking the series of Poisson observations, and making accurate representations of the foreground and background. It is especially important to notice that including the foreground and background dynamics allows for the foreground image to become clear. Without incorporating dynamics and regularization, the additional foreground image would simply be the transient errors of the background estimation. By including the dynamics in the optimization process, the systematic difference of the background estimate can be found to be the foreground object.

4 Conclusions

The research supported by this grant resulted in several innovative methods and supporting theory for photon-limited sensing and surveillance. We have developed practical methods representative of the current state-of-the-art with online code actively used by the signal processing community. We have also developed theory that highlights the challenges of photon-limited imaging in compressed sensing contexts and described the potential benefits of using conventional imagers rather can compressive imagers. In addition, we have developed novel theory that characterizes the performance of online learning methods which learn and exploit underlying dynamics. These fundamental performance bounds relating to reconstruction accuracy and regret bounds associated with sequential processing of video frames guided the development of fast and novel computational techniques, and set the stage for further advances in future work.

References

[1] F. J. Anscombe. The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 35: 246–254, 1948.

[2] A. Bain and D. Crisan. *Fundamentals of Stochastic Filtering*. Springer, 2009.

[3] L. Balzano, R. D. Nowak, and B. Recht. Online identification and tracking of subspaces from highly incomplete information. In *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pages 704–711. IEEE, 2010.

[4] A. Beck and M. Teboulle. Mirror descent and nonlinear projected subgradient methods for convex programming. *Operations Research Letters*, 31:167–175, 2003.

[5] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions Image Processing*, 18(11):2419–34, 2009.

[6] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J-B. Sibarita, and J. Salamero. Patch-based nonlocal functional for denoising fluorescence microscopy image sequences. *IEEE Trans. Med. Imag.*, 29(2): 442–454, 2010.

[7] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning and Games*. Cambridge University Press, New York, 2006.

[8] T. Chan and J. Shen. *Image Processing And Analysis: Variational, PDE, Wavelet, And Stochastic Methods*. Society for Industrial and Applied Mathematics, 2005.

[9] M. Fisz. The limiting distribution of a function of two independent random variables and its statistical application. *Colloquium Mathematicum*, 3:138–146, 1955.

[10] P. Fryźlewicz and G. P. Nason. Poisson intensity estimation using wavelets and the Fisz transformation. Technical report, Department of Mathematics, University of Bristol, United Kingdom, 2001.

[11] E. Hall and R. Willett. Dynamical models and tracking regret in online convex programming. *arXiv:1301.1254*, 2013. To appear in *Proc. ICML*.

[12] M. Mäkitalo and A. Foi. Optimal inversion of the Anscombe transformation in low-count Poisson image denoising. *IEEE Trans. Image Process.*, 20(1):99–109, 2011.

[13] N. Merhav and M. Feder. Universal prediction. *IEEE Trans. Info. Th.*, 44(6):2124–2147, October 1998.

[14] A. S. Nemirovsky and D. B. Yudin. *Problem complexity and method efficiency in optimization*. John Wiley & Sons, New York, 1983.

[15] M. Piccardi. Background subtraction techniques: a review. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 4, pages 3099 – 3104, Oct. 2004. doi: 10.1109/ICSMC.2004.1400815.

[16] M. Raginsky, R. Willett, Z. Harmany, and R. Marcia. Compressed sensing performance bounds under Poisson noise. *IEEE Transactions on Signal Processing*, 58(8):3990–4002, 2010. arXiv:0910.5146.

[17] J. Salmon, Z. Harmany, C. Deledalle, and R. Willett. Poisson noise reduction with non-local PCA. *Journal of Mathematical Imaging and Vision*, 48(2):279–294, 2014. arXiv:1206:0338.

[18] D. Snyder. *Random Point Processes*. Wiley-Interscience, New York, NY, 1975.

[19] C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2, 1999. doi: 10.1109/CVPR.1999.784637.

[20] Y. Theodor and U. Shaked. Robust discrete-time minimum-variance filtering. *IEEE Trans. Sig. Proc.*, 44(2):181–189, 1996.

[21] L. Xie, Y. C. Soh, and C. E. de Souza. Robust Kalman filtering for uncertain discrete-time systems. *IEEE Trans. Autom. Control*, 39:1310–1314, 1994.

[22] B. Zhang, J. Fadili, and J-L. Starck. Wavelets, ridgelets, and curvelets for Poisson noise removal. *IEEE Trans. Image Process.*, 17(7):1093–1108, 2008.

[23] M. Zinkevich. Online convex programming and generalized infinitesimal gradient descent. In *Proc. Int. Conf. on Machine Learning (ICML)*, pages 928–936, 2003.